A Genetic Bottleneck of Mitochondrial DNA During Human Lymphocyte Development

Zhongjie Tang,[†,1] Zhaolian Lu,[†,2] Baizhen Chen,[1] Weixing Zhang,[1] Howard Y. Chang,[*,3,4] Zheng Hu,[*,2] and Jin Xu[*,1]

[1]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou, China

[2]CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[3]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA

[4]Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

***Corresponding authors:** E-mails: howchang@stanford.edu; zheng.hu@siat.ac.cn; xujin7@mail.sysu.edu.cn.

[†]These authors contributed equally to the work.

Associate editor: Bing Su

Abstract

Mitochondria are essential organelles in eukaryotic cells that provide critical support for energetic and metabolic homeostasis. Although the elimination of pathogenic mitochondrial DNA (mtDNA) mutations in somatic cells has been observed, the mechanisms to maintain proper functions despite their mtDNA mutation load are poorly understood. In this study, we analyzed somatic mtDNA mutations in more than 30,000 single human peripheral and bone marrow mononuclear cells. We observed a significant overrepresentation of homoplasmic mtDNA mutations in B, T, and natural killer (NK) lymphocytes. Intriguingly, their overall mutational burden was lower than that in hematopoietic progenitors and myeloid cells. This characteristic mtDNA mutational landscape indicates a genetic bottleneck during lymphoid development, as confirmed with single-cell datasets from multiple platforms and individuals. We further demonstrated that mtDNA replication lags behind cell proliferation in both pro-B and pre-B progenitor cells, thus likely causing the genetic bottleneck by diluting mtDNA copies per cell. Through computational simulations and approximate Bayesian computation (ABC), we recapitulated this lymphocyte-specific mutational landscape and estimated the minimal mtDNA copies as <30 in T, B, and NK lineages. Our integrative analysis revealed a novel process of a lymphoid-specific mtDNA genetic bottleneck, thus illuminating a potential mechanism used by highly metabolically active immune cells to limit their mtDNA mutation load.

Key words: **mitochondrial DNA, somatic genetic bottleneck, lymphocyte, scATAC-seq.**

Introduction

Mitochondria are essential cellular organelles that orchestrated a wide range of biological processes, including energy production and oxidative metabolism (Chandel 2014). Immune cells have high and dynamic metabolic requirements for their immune functions in healthy and diseased states, which rely on the mitochondrial biogenesis. For instance, the mitochondrial dynamics controls T cell fate through metabolic programming (Buck et al. 2016). Extensive mitochondrial remodeling, including alterations in morphology and increases in the copy numbers of mitochondrial DNA (mtDNA) during T cell activation has been reported (Tan et al. 2017; Mimitou et al. 2021). The metabolic alterations in immune cells have been observed not only in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection (Lee et al. 2022), but also in the antitumor activity of tumor-infiltrating T lymphocytes (Yu et al. 2020). The mitochondrial translation is required for sustained killing by cytotoxic T cells and tumor-infiltrating T lymphocytes with depolarized mitochondria display characteristics of terminally exhaustion (Yu et al. 2020; Lisci et al. 2021).

The mitochondria function relies on the mtDNA, which encodes 13 core peptide subunits of the oxidative phosphorylation system and 24 RNAs involved in intramitochondrial protein synthesis (Andersion et al. 1981, Andrews et al. 1999). Typically, hundreds to thousands of copies of mtDNA molecules are present in each cell, and the germline mtDNA is predominantly maternally inherited and does not undergo recombination (Hagström et al. 2014). MtDNA accumulates mutations at a rate that is 5–10 times higher per site than that in the nuclear genome, because of the lack of DNA repair systems and frequent contact with mutagenic reactive oxygen species (Kazak et al. 2012; Lagouge and Larsson 2013; Scheibye-Knudsen et al. 2015). More than 500 pathogenic mtDNA mutations have been identified as genetic defects causing various human diseases (Ye et al. 2014). According to the theory known as "Muller's ratchet," continual

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6572... by guest on 16 May 2022

Article

accumulation of deleterious mutations in the absence of purifying selection leads to a decline in population fitness and will ultimately result in mutational meltdown (Felsenstein 1974). To avoid this outcome, the animal germline has evolved a mitochondrial genetic bottleneck, wherein only a small subset of mtDNA is transmitted to the next generation, thus resulting in substantial removal of deleterious mutations (Koehler et al. 1991; Ghosh et al. 1996; Jenuth et al. 1996).

A recent case report had shown that the pathogenic mutation 3243A/G, the cause of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episode (Walker et al. 2020), was remarkably purified in lymphoid immune cells, particularly in T cells, as compared with other blood cells from peripheral blood mononuclear cells (PBMCs). A rare blood sample from a patient with Pearson's syndrome in an acute anemic state caused by mtDNA mutations, spontaneously recovered with clearance of erythroid mitochondrial (Ahlqvist et al. 2015). These observations rose the questions on how pathogenic mtDNA mutations are inherited in somatic cells and how they affect the immune system.

To answer these questions, understanding of the clonal dynamics of mtDNA in the development of somatic cell lineages is necessary; however, systematical and quantitative studies are lacking due to the technical limitations of detecting heteroplasmic mutations in single cells. We and others have recently developed a single-cell lineage tracing method leveraging the somatic mtDNA mutations detected in single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq) and/or RNA-seq (scRNA-seq) data (Ludwig et al. 2019; Xu et al. 2019). Using this method, we systematically investigated the mtDNA mutation landscape in 30,000 human single peripheral and bone marrow mononuclear cells (BMMCs). We revealed a novel process of genetic bottleneck in lymphoid lineage development and quantified the size of the genetic bottleneck with the approximate Bayesian computation (ABC) method. Together, our results demonstrated a novel mechanism that may strengthen purifying selection and consequently enable better quality control of their mitochondrial genomes, owing to the metabolic needs for immune responses.

Results

Somatic Mutational Landscape of mtDNA at Single-Cell Resolution

In this study, we focused on the human hematopoietic system, whose cellular differentiation lineages have been well documented. We first identified somatic mtDNA mutations in a previously reported mitochondrial scATAC-seq (or mtscATAC-seq) dataset including more than 20,000 blood cells from a healthy 47-year-old individual (fig. 1*a* and *b*, supplementary fig. S1*a*, Supplementary Material online, see Materials and Methods) (Lareau et al. 2021). We summarized the numbers of mutations and the variant allele frequency (VAF, also referred to as mtDNA heteroplasmic ratio) in each cell to compare the VAF distribution in a population of different cell types. Interestingly, cells of the mature lymphocyte lineages—specifically B, T, and natural killer (NK) cells—had a significantly lower mtDNA mutational burden than hematopoietic progenitor cells, including hematopoietic stem cells (HSCs), multipotent progenitors (MPPs), lymphoid-primed MPPs (LMPPs), common lymphoid progenitors (CLPs), common myeloid progenitors (CMPs), and granulocyte–macrophage progenitors (GMPs) (fig. 1*c*, supplementary fig. S1*b*, Supplementary Material online, Wilcoxon test, $P < 2.2e^{-16}$). The mtDNA mutational burden was also lower in lymphocytes than the myeloid and erythroid lineages (fig. 1*c*, Wilcoxon test, $P < 2.2e^{-16}$). As anticipated, most somatic mtDNA mutations were detected at low VAF in individual cells in all cell types (fig. 1*d*). However, the distribution of somatic homoplasmic mutations (i.e., those with VAF of ~1) varied substantially among the different cell types. For instance, progenitor cells, including HSCs, MPPs, LMPPs, CLPs, CMPs, and GMPs exhibited a typical monotonic decline in the number of mutations with increasing VAF (fig. 1*d*). Although this pattern was also seen in both the myeloid and erythroid lineages (e.g., monocytes and erythrocytes), we observed an unanticipated increase in the number of somatic homoplasmic mutations in B, T, and NK cells (fig. 1*d*).

In addition to the mtscATAC-seq dataset from Lareau et al., we analyzed another mtscATAC-seq dataset of 10,327 BMMCs from an independent healthy donor (Mimitou et al. 2021) (fig. 2*a*). Lymphocytes in BMMCs also carried a lower mtDNA mutational burden with a characteristic overrepresentation of somatic homoplasmic mutations (fig. 2*b*, supplementary fig. S1*c* and *d*, Supplementary Material online). Furthermore, to exclude the potential bias of mutations arising at different development time points, we summarized the allele frequency of individual mtDNA mutations shared between lymphoid and myeloid cell lineages. The allele frequency spectrum of the variant 16356T/C and 10304G/A demonstrated that the proportion of cells with homoplasmic variants in B (22% and 33%), T (48% and 40%), and NK (41% and 44%) cells was significantly higher than that in monocytes (10% and 6%) (fig. 2*c*–*f*). In fact, these lymphocyte-specific characteristics were also verified by additional scATAC-seq or scRNA-seq data from seven independent individuals. And the observations were aslo independent of sequencing depth (supplementary figs. S1*e* and S2–S4, Supplementary Material online), thus indicating a general and unique process of mtDNA inheritance in lymphocyte development.

Asynchronous Replication of Mitochondrial and Nuclear Genomes During B Cell Development

To examine whether the distinct VAF distribution between lymphoid cells and myeloid/erythroid cells was due to the variation in mtDNA copy number per cell, we estimated the

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

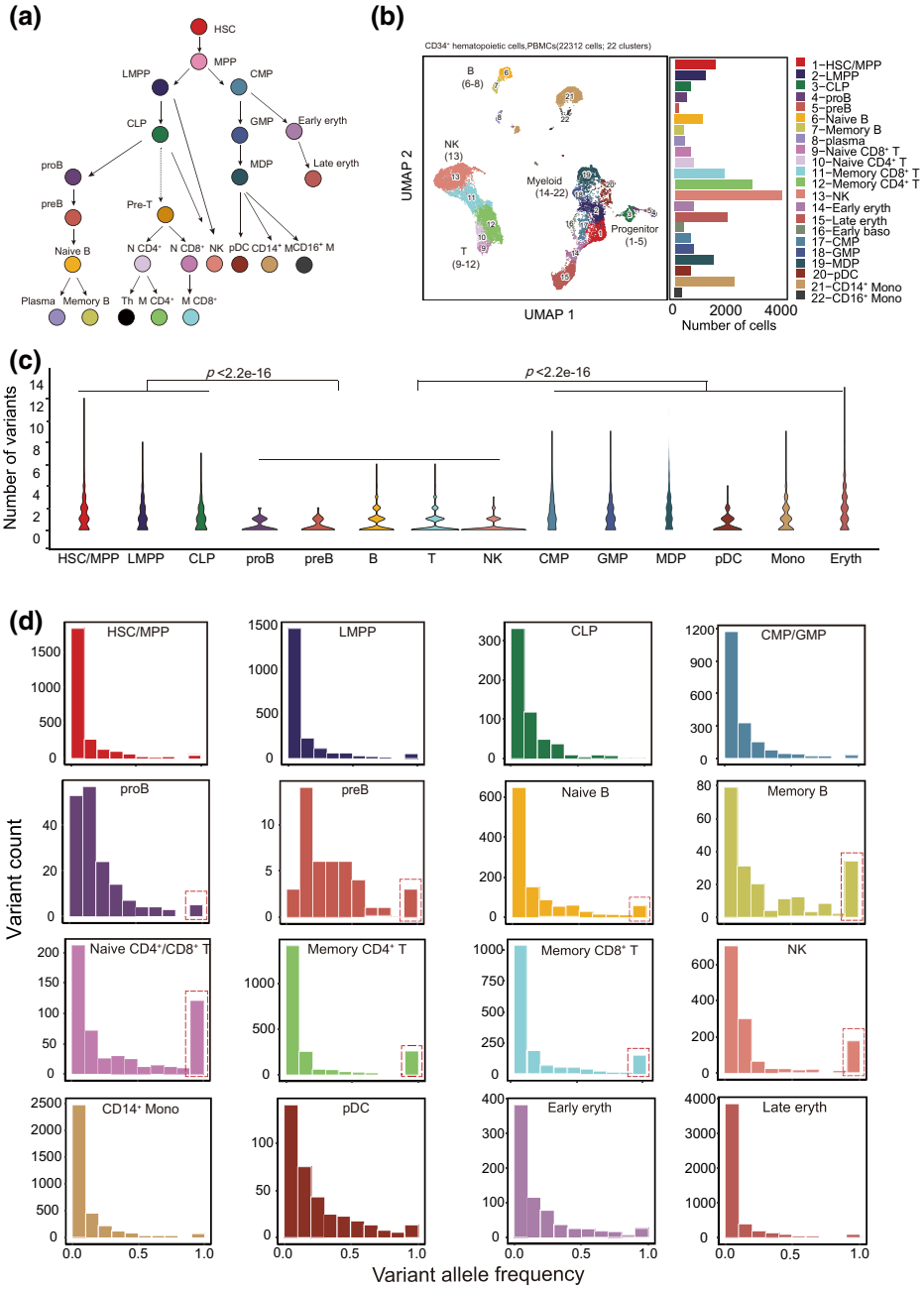

Fig. 1. Somatic mutations in the mtDNA of PBMCs. (*a*) Schematic of human hematopoietic differentiation and lineage commitment. HSC, hematopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte–monocyte progenitor; MDP, monocyte-dendritic cell progenitor; N CD4, naive $CD4^+$ T cell; N CD8, naive $CD8^+$ T cell; M CD4, memory $CD4^+$ T cell; M CD8, memory $CD8^+$ T cell; Th, T helper cell; NK, natural killer cell; pDC, plasmacytoid dendritic cell; Eryth, erythrocyte. (*b*) UMAP projection of 22,312 CD34 + hematopoietic cells and PBMCs with mtscATAC-seq data. Dots represent individual cells, which are colored according to cluster identity. The bar plot indicates the number of cells in each cluster (labeled at right). (*c*) Violin plot showing the number of somatic mtDNA variants per cell for various cell types; *P*-values, two-sided Wilcoxon rank-sum test. (*d*) The VAF distribution of somatic mtDNA mutations across different cell types. Somatic homoplasmic mutations (VAF of ~1) identified in the lymphoid lineage are highlighted with a red box.

relative number of mtDNA copies in each cell type according to the fraction of sequencing reads mapped to the mitochondrial genome relative to the total number of reads in each cell (fig. 3*a*, supplementary fig. S5*a*, Supplementary Material online). Although mature lymphocytes and progenitor cells had similar mtDNA copy numbers, pro-B, and pre-B cells—the earliest lineage-committed cells in B cell development—exhibited a significantly lower number of mtDNA copies (Wilcoxon test, pro-B/pre-B vs. HSC/MPP, $P < 2.2e^{-16}$; pro-B/pre-B vs. B, $P < 2.2e^{-16}$). Of note, the CLPs also showed significantly fewer mtDNA copies than earlier progenitors did (Wilcoxon test, CLP vs. HSC/MPP, $P < 2.2e^{-16}$; CLP vs. LMPP, $P < 2.2e^{-16}$), thus indicating a marked mtDNA copy number decrease in early lymphocyte development. Therefore, we hypothesized that the characteristic mutational spectra in lymphocyte mtDNA (figs. 1*c*, *d* and 2*b*) might have resulted from a mitochondrial genetic bottleneck.

To address this possibility, we examined the mtDNA replication machinery to gain insight into the regulation of mtDNA copy number along the lymphocyte differentiation trajectory. We focused on the B cell lineage, because T cells mature in the thymus and their progenitors, pre-T cells, are not present in PBMCs. DNA polymerase γ is the only known mtDNA polymerase in animals (Radsak and Schutz 1978), has both a catalytic (*POLG*) and a binding subunit (*POLG2*), and it catalyzes the polymerization of deoxyribonucleotides. High levels of DNA polymerase γ activity have been detected in the S and G2 cell cycle phases, thereby maintaining stable numbers of mtDNA during cell

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

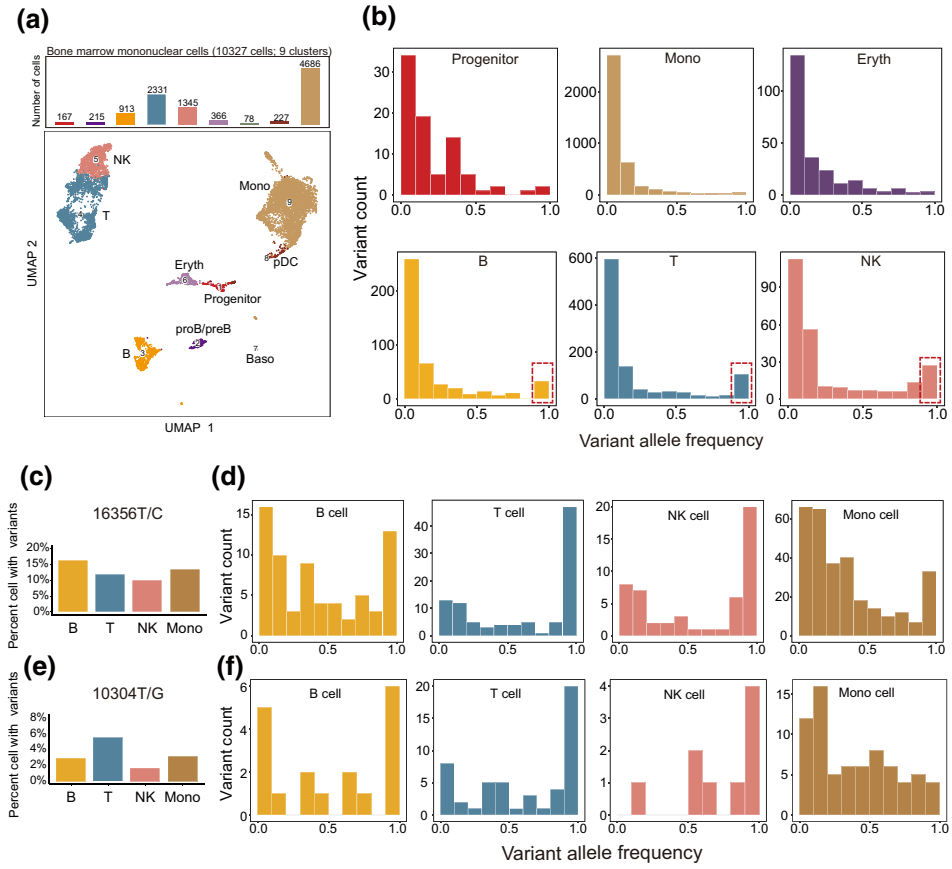

Fig. 2. Somatic mutations in the mtDNA of BMMCs. (*a*) UMAP projection of 10,327 mononuclear cells from bone marrow with mtscATAC-seq data. Dots represent individual cells, which are colored according to cluster identify and cell types. (*b*) The VAF distribution of somatic mtDNA mutations across different cell types in BMMCs. Somatic homoplasmic mutations (VAF ~1) identified in the lymphoid lineage are highlighted with a red box. (*c*) Distribution of the observed mtDNA variant 16356T/C in cells among lymphoid and myeloid cell lineages. (*d*) The VAF distribution of the somatic mtDNA mutation 16356T/C across different cell types. (*e*) Distribution of the observed mtDNA variant 10304G/A in cells among lymphoid and myeloid cell lineages. (*f*) The VAF distribution of the somatic mtDNA mutation 10304G/A across different cell types.

division (Radsak and Schutz 1978; Chatre and Ricchetti 2013; Sasaki et al. 2017). To determine whether the expression of DNA polymerase γ increases with cell proliferation during B cell development, we projected the developmental trajectory of cell subpopulations from HSCs to mature B cells via a pseudotime analysis with scRNA-seq data (fig. 3*b* and supplementary fig. S5*b*, Supplementary Material online). We observed up-regulation of G1/S phase-specific genes (such as DNA polymerase δ, *POLD1*–3) in both pro-B and pre-B cell populations, thus suggesting high activation of cell proliferation in these cell types (fig. 3*c* and *d*, supplementary fig. S5*c*–*e*, Supplementary Material online). In contrast, the expression of DNA polymerase γ was not coupled with cell proliferation (fig. 3*e* and *f*, supplementary fig. S5*c*–*e*, Supplementary Material online). Unexpectedly, the expression of the DNA polymerase γ binding subunit (*POLG2*) was significantly diminished in the highly proliferative pro-B and pre-B cell subpopulations (fig. 3*g*). Together, these results implied a genetic bottleneck during B cell development which might have resulted from limited replication of mtDNA, thus diluting the mtDNA copy number throughout cell division.

Quantification of the mtDNA Genetic Bottleneck by Computational Modeling

To test whether the genetic bottleneck was achieved by the dilution of mtDNA and quantify the extent of the mitochondrial genetic bottleneck, we developed a computational model of an mtDNA dilution process based on population genetics theory (fig. 4*a*). In this model, we assumed the starting mtDNA copy number is N_0 (~500 copies per cell estimated by O'Hara et al. 2019) in LMPP cells and only a proportion of mtDNA molecules (denoted by α) replicate during each cell cycle. This process of dilution starts from LMPPs and continues for T_d cell generations until the mtDNA copy number recovers to the starting level (~500 copies). Using the ABC method (see Materials and Methods), we estimated the model parameters for B, T, and NK cell populations by using a constant mtDNA mutation rate of 10^{-7} per site per cell division (fig. 4*b*, supplementary fig. S6*a*, Supplementary Material online) (Coller et al. 2001). The model estimated the minimal mtDNA copy number to be 21 (95% confidence interval [CI] = 13–56), 13 (95% CI = 12–19), and 14 (95% CI = 12–29) in each B, T, and NK cell, respectively. These values were 20–40-fold lower than the normal mtDNA levels. The VAF distribution simulated with these parameter estimations recapitulated the observed data, showing a characteristic overrepresentation of somatic homoplasmic mutations (i.e., VAF of ~1) and a diminished overall mutational burden (fig. 4*c*, supplementary fig. S6*b* and *c*, Supplementary Material online). Assuming a higher starting mtDNA copy numbers (~1,000 per cell) resulted in very similar estimations regarding the bottleneck size (supplementary fig. S7*a* and *b*, Supplementary Material online). Therefore, this pattern can not be achieved by

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

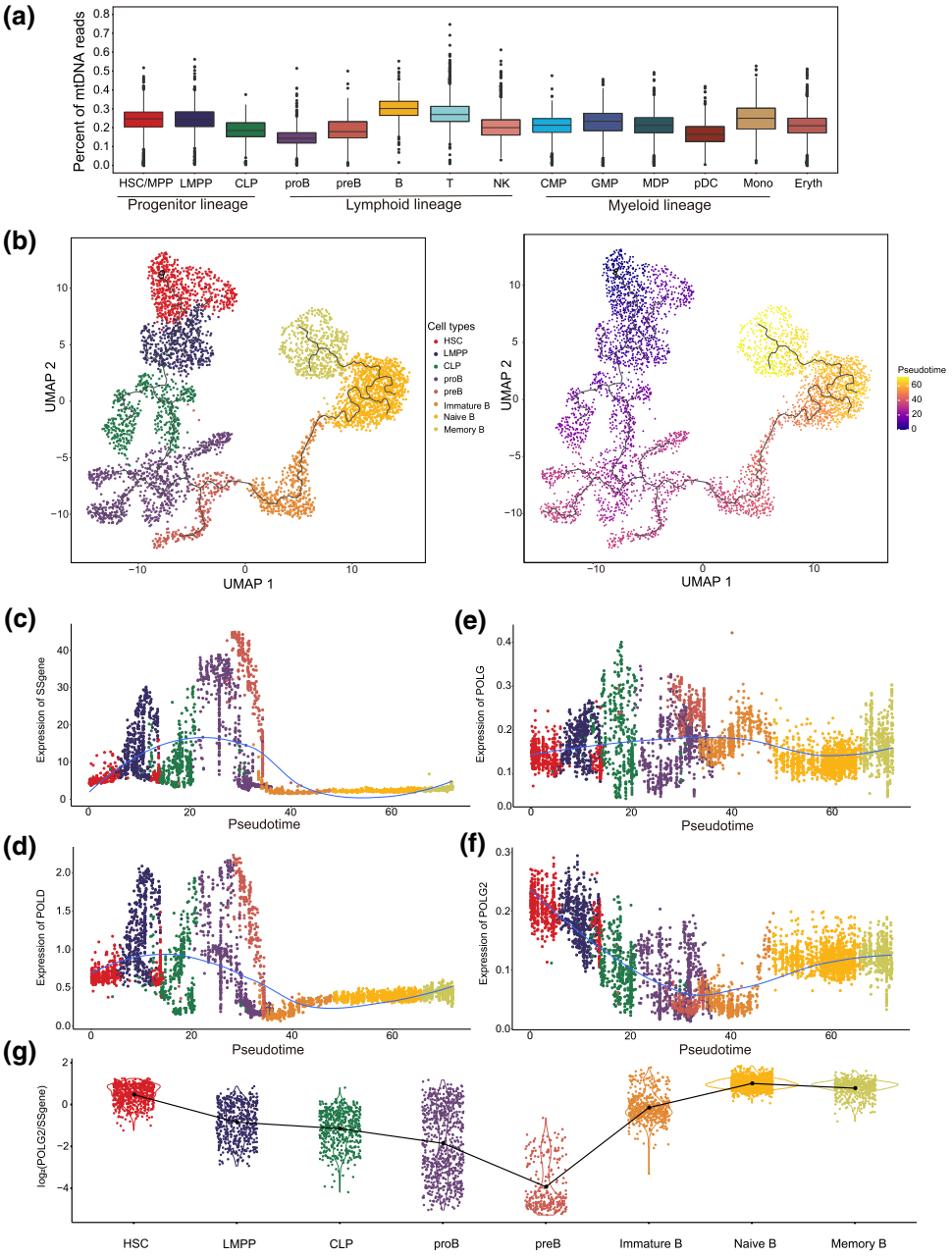

Fig. 3. Replication of mtDNA during B cell development. (*a*) The relative number of mtDNA copies was determined according to the proportion (sequencing reads mapped to mitochondrial genome divided by the total number of reads) in each cell, as identified from the scATAC-seq dataset. (*b*) Pseudotime trajectory of B cell differentiation from HSCs by using single-cell RNA-seq data generated from PBMCs ($n = 4692$ cells). Colors denote different cell types (top) and developmental stages (bottom) defined by pseudotime. The solid line represents the fitted trajectory across pseudotime. (*c–f*) Kinetic plots showing the expression of (c) 39 G1/S phase-specific genes (SSgene), (*d*) nuclear DNA polymerase δ (*POLD1–3*), (*e*) mtDNA polymerase γ (*POLG*), and (*f*) the binding subunit of mitochondrial DNA polymerase γ (*POLG2*) along the B-cell developmental trajectory. (*g*) Violin plot showing the ratio of *POLG2* expression to the mean expression of all G1/S phase-specific genes in each cell associated with B-cell development. The broken line represents the change trend of the mean ratio across different cell types.

random genetic drift with a constant number of mtDNA copies during the lymphocyte development.

The Consequences of the mtDNA Genetic Bottleneck

Collectively, our integrative genomic data analysis and computational modeling demonstrated the existence of a stringent mtDNA genetic bottleneck that resulted from replicative dilution during lymphocyte development. This mechanism strengthens the genetic drift toward a lower mtDNA mutational burden and lower genetic diversity within each cell. We wondered whether the genetic bottleneck during lymphocyte development might have the same purifying selection effects as those in the germline. We, therefore, examined the VAF distribution in various genomic regions (loop, tRNA, rRNA, and coding) or mutation types (synonymous [SY] and nonsynonymous [NS]). We found loop and SY regions had higher proportions in somatic homoplasmic mutations than in tRNA, rRNA, and NS regions in datasets (fig. 5*a*, supplementary fig. S8, Supplementary Material online). This result indicated that the mutations in coding regions were more likely to be selected against, when they reached to a high frequency. We next examined the *dN*/*dS* ratio (the ratio of the normalized number of NS substitutions—*dN* to the normalized number of SY substitutions—*dS*) (supplementary fig. S9, Supplementary Material online) according to heteroplasmic categories. The calculated *dN*/*dS*

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

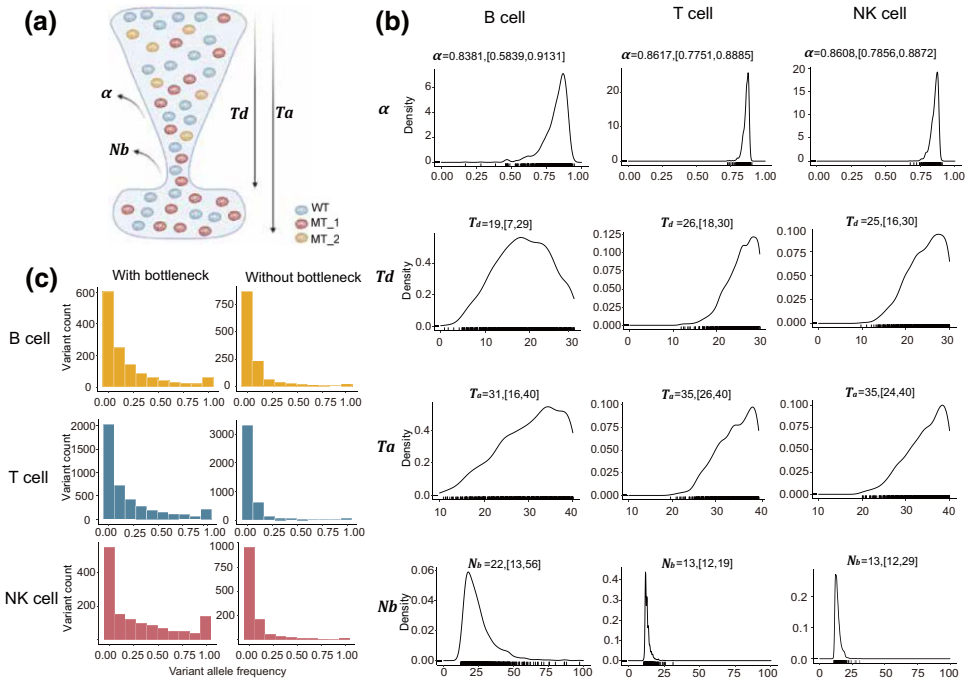

FIG. 4. A dilution model of mitochondrial genetic bottleneck. (*a*) Schematic illustration of the dilution model of the mitochondrial genetic bottleneck. In this model, only a fraction of mtDNA molecules (denoted by α) replicate at each cell division. After T_d cell divisions from the LMPP stage, the number of mtDNA copies in each lymphocyte subtype (B, T, and NK cells) rapidly recovers to the baseline level (~500 per cell). N_b denotes the minimal number of mtDNA copies that can be computed as equation (1). The total number of cell divisions required for the transition from LMPP to mature lymphocyte is denoted as T_a. (*b*) The distribution of model parameters inferred by the ABC algorithm. The mean and 95% CI of each parameter estimation are shown. (*c*) Simulations based on the dilution model of mitochondrial genetic bottleneck with the ABC-estimated parameter values recapitulated the lymphocyte-specific overrepresentation of somatic homoplasmic mutations and the lower mutation burden (supplementary fig. S6*b*, Supplementary Material online). The left and right panels represent the simulations with and without mitochondrial genetic bottleneck, respectively. The average of 100 simulations carried out for each model is as shown. The results of each iteration are shown in supplementary fig. S6*c*, Supplementary Material online.

ratios revealed a pattern of generally neutral evolution (i.e., $dN/dS \sim 1$) in all categories in most of the cases examined (supplementary fig. S9, Supplementary Material online). The dN/dS ratio might not be sensitive in testing selection in somatic cells, owing to the linkage of whole mitochondrial genome with strong Hill–Robertson interference, thus leading to a pattern of quasi-neutrality, as in cancer evolution (Chen et al. 2019). Therefore, we verified individual mutation sites to identify the signals of purifying selection; indeed, we observed several mutations that were specifically eliminated in lymphocytes but not the myeloid lineage (fig. 5*b*). For example, the mutations, 2636G/A, and 3209A/G, underwent the most profound decrease in prevalence (fig. 5*c*) in lymphocytes. Intriguingly, these two sites are all located at MT-RNR2, which encodes 16S rRNA and Humanin, a peptide protective against multiple mitochondrial diseases (fig. 5*d*) (Lee et al. 2013). Furthermore, we queried MITOMAP, a human mitochondrial genome database, and found that mtDNA variants reported on MT-RNR2 were highly associated with sepsis (Wilcoxon test, $P < 2.2e^{-16}$, fig. 5*e*) (Lott et al. 2013; Park et al. 2021); therefore, MT-RNR2 may play important roles in immune functions protecting against infections. These data indicated that purifying selection in lymphocytes indeed occurs for specific mtDNA mutation sites.

Discussion

Collectively, our results indicated an unanticipated lower mutational burden and accumulation of homoplasmic mtDNA mutations in lymphocytes, thus indicating a stringent genetic bottleneck and purifying selection of mtDNA. Gene expression data and computational modeling suggested a dilution process, based on the rate of mtDNA replication relative to the nuclear genome.

The single-cell data derived from PBMCs could not capture the full developmental trajectory of T cells because pre-T cells develop in the thymus. Our single-cell data and computational inference indicated that the genetic bottleneck in T and NK cells might be as stringent as that in B cells (figs. 1*d*, 2*a*, and 4*b*). On the basis of our observations and simulations, we hypothesized that the regulation of the lymphocyte-specific genetic bottleneck may start from the CLP stage, instead of subsequent lineage commitment for B, T, and NK cells. The effect of this regulation is likely to be enhanced via the active proliferation of progenitor cells. During lymphocyte development,

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

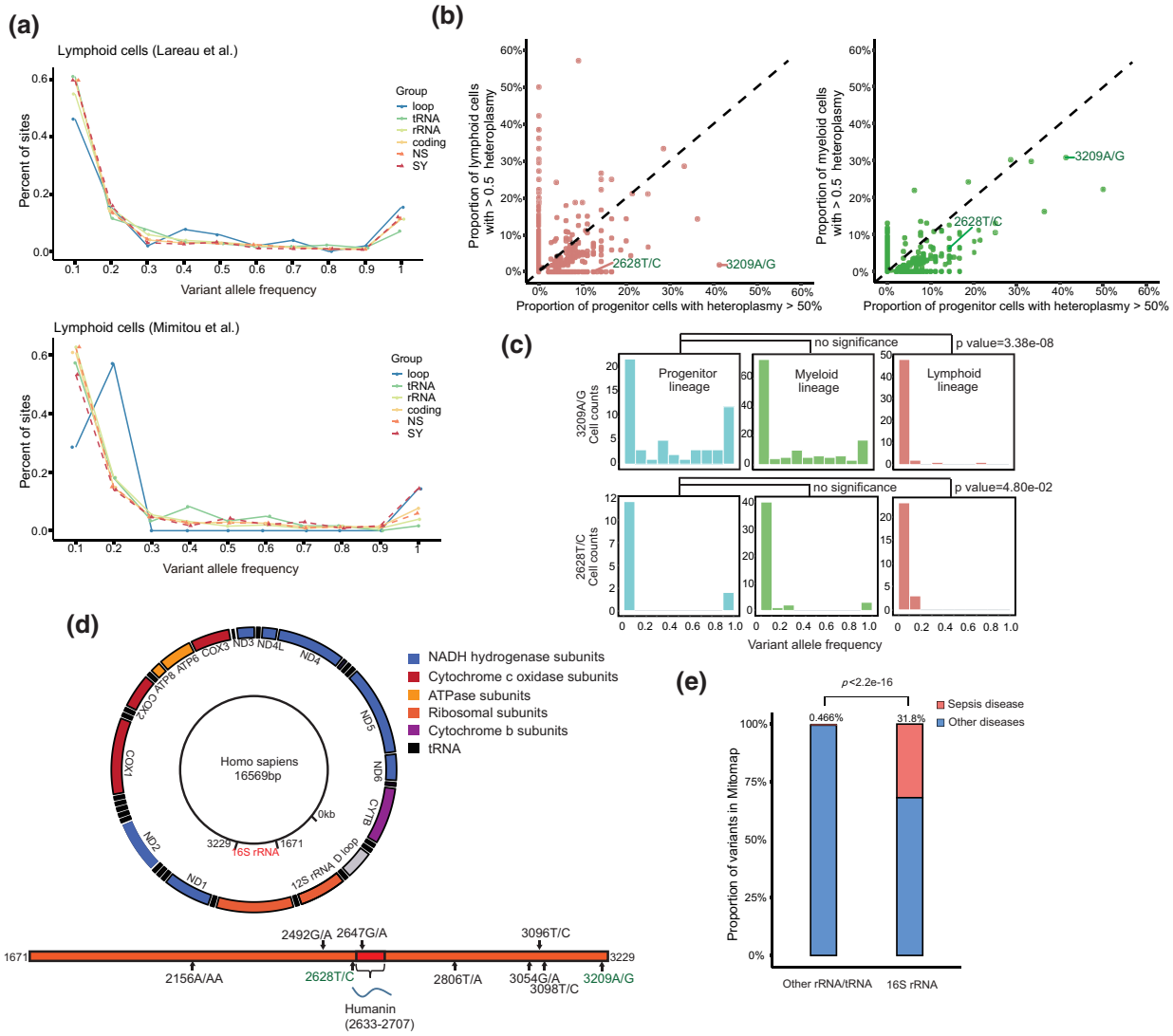

Fig. 5. Elimination of specific mtDNA variants in lymphocyte. (*a*) Distribution of the VAF for mutations in different mtDNA genomic regions in lymphoid cells. The color code corresponds to mtDNA genomic regions or mutation types, annotated as loop, tRNA, rRNA, coding (coding region), NS (nonsynonymous), and SY (synonymous). (*b*) Scatter plot showing the percentage of cells with dominant mtDNA mutations (VAF > 50% in a single cell). Results from progenitor cells (HSC, MPP, and LMPP) compared with cells from lymphoid (B, T, or NK cells) or myeloid lineages (right) are shown. (*c*) The distribution of VAF for two individual sites (3209A/G, 2636G/A) in progenitor, myeloid, and lymphoid cells, respectively. The *P*-values shown were determined with the χ^2 test. (*d*) The location of 16S RNA (MT-RNR2) on the mitochondrial genome and the locations of sepsis associated variants on MT-RNR2, reported in MITOMAP (in black), or those specifically eliminated in lymphocytes (in green). (*e*) The proportion of mtDNA variants associated with sepsis disease in 16S RNA vs. other rRNA/tRNA genes in the mitochondrial genome.

multipotent T and B progenitor cells undergo a series of maturation steps that include positive selection for functional T-cell receptors or immunoglobulins and negative selection to eliminate cells with a high affinity toward self-associated peptides or antigens (Starr et al. 2003). Only a small proportion of T lymphoid cells will survive after the negative and positive selections. Moreover, mitochondrial function is important for T cell development and their functional activation (Buck et al. 2016; Chao et al. 2017). The metabolic responses characteristic of lymphocytes development and activation are both well-regulated at the transcriptional and posttranscriptional levels (Maciver et al. 2013). For example, several groups have shown that T or B cell activation leads to mitochondrial remodeling and dramatic shifts in cell metabolism, in the course of their role in eliminating pathogens (Sena et al. 2013; West and Shadel 2017; Angajala et al. 2018; Waters et al. 2018; Lee et al. 2022). Furthermore, the selection against the pathogenic mutations 3243 was stronger in T cells than in B and NK cells, as shown by Walker et al. (2020). Further systematic study of T precursors in thymus will provide further insight into how genetic bottlenecks occur during T cell development.

On the basis of the scRNA-seq data, we found higher expression of Drp1 and LC3 genes in pro/pre-B cells than the other progenitor cells, thus implying that mitochondria at

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

this stage may undergo fission or mitophagy process (Lemasters 2005). Further experimental evidence will be helpful to determine whether mitophagy of mitochondria is also involved in the genetic bottleneck process in the progenitor cell stage. If so, the selection of mtDNA mutations might also happen in mitochondrial organelles level.

Our computational framework modeled a single-cell lineage and quantitatively measured the mtDNA bottleneck size after the continual dilution process. Because B/T/NK cell development and maturation are complicated process involving a series of clonal selections on cellular levels, such as positive and negative selection of pre-B/pre-T cells in the primary lymphoid organs (Starr et al. 2003), we were unlikely to model all those selection processes on specific cell types within our model. Therefore, to make our model simple and robust, we only considered a neutral process of continual dilution with mtDNA copy number changes at the single-cell level. Furthermore, the mutation rate in our model is sensitive to the detection of low-frequency mutations. Ultradeep sequencing of the mitochondrial genome through other single-cell mtDNA sequencing methods (Guo et al. 2020) would aid in estimating the mutation rate accurately and would further facilitate the estimation of selection strength.

Our newly discovered somatic mtDNA bottleneck within the lymphoid lineage may play a role in the quality control of mitochondrial genomes, in parallel to the selection of immunoreceptor genes in the nuclear genome. Thus, a robust population of mtDNA may be crucial for lymphocyte-mediated immune responses. This mtDNA genetic bottleneck may be one of several potential mechanisms involved in the regulation of mitochondrial genome in different lineages. Currently, there is not available for high-quality mtscATAC datasets in many different cell types, it is unknown whether the stringent mtDNA genetic bottleneck is a lymphocyte-specific phenomenon or also prevalent in other cell types. Hence, the prevalence, causes, and the consequences of the somatic mtDNA genetic bottleneck require extensive exploration in future studies. The understanding of the mtDNA quality control mechanism may ultimately provide new insights into immune degeneration and related diseases, and contribute to the treatments based on engineered immune cells, such as chimeric antigens receptor-T cell therapy.

Materials and Methods

Data Collection

The mtscATAC-seq dataset generated through evaluation of hematopoietic and PBMCs was retrieved from a recent study evaluating samples from a healthy 47-year-old donor (Lareau et al. 2021). The mtscATAC-seq dataset from human bone marrow from a healthy 25-year-old donor was obtained from Mimitou et al. (2021). The scATAC-seq data from $CD4^+$ T cells were obtained from the study published by Satpathy et al (2018). The scATAC-seq dataset for HSCs, MPPs, LMPPs, CLPs, CMPs, GMPs, and plasmacytoid dendritic cells (pDCs) derived from $CD34^+$ bone marrow was obtained from Buenrostro et al (2018) (supplementary fig. S2, Supplementary Material online). The scRNA-seq dataset generated from an evaluation of healthy $CD34^+$ PBMCs, BMMCs, and total PBMCs was downloaded from a study published by Granja et al. (2019). These datasets were used to analyze mtDNA replication and gene transcription. The scRNA-seq dataset of 70 effector memory T cells (Tem cells), 70 central memory T cells (Tcm cells), and 142 $CD4^+$ regulatory T cells (Treg cells) from healthy human colon tissue were downloaded from Array Express (E-MTAB-6072) (Miragaia et al. 2019). Detailed information on data resources is provided in supplementary table S1, Supplementary Material online.

Single-Cell (sc)ATAC-seq Data Preprocessing and Annotation of the Cell Populations

Raw data from GSE142745 were processed with Cell Ranger ATAC (version 2.0.3; 10× Genomics, https://www.10xgenomics.com/products/single-cell-atac) with default parameters. Reads were aligned to the reference hg19 human genome (https://support.10xgenomics.com/single-cell-atac/software/downloads/latest). In each cell, 40% of fragments overlapping a compendium of DNase hypersensitivity peaks and 1,000 unique nuclear fragments were filtered. From the output of the Cell Ranger Software calls, we performed a computational annotation of the cell types on the basis of chromatin accessibility. Clustering and gene activity scores were determined through standard processing via ArchR (Granja et al. 2021). Clustering was performed with the "addClusters" and "addUMAP" functions (resolution = 0.8, neighbors = 10, minDist = 0.1). To identify marker genes according to gene scores, we used the "getMarkerFeatures" function with useMatrix "GeneScoreMatrix" and generated a reproducible peak set in ArchR by using the "addReproduciblePeakSet" function. By default, ArchR attempts to identify peaks by using the MACS2 algorithm (Zhang et al. 2008). Because common cell markers are sometimes not suitable for classification with "GeneScoreMatrix," we used enhancer accessibility to define the cell type. For example, we identified myeloid cells according to the unique accessibility of enhancers at +85 and +87 kb in the interferon regulatory factor (*IRF8*) locus. pDCs were identified on the basis of the unique accessibility of +54 and +56 kb enhancers, as described by Satpathy et al (2018). Furthermore, to label scATAC-seq clusters with scRNA-seq information, we used the "addGeneIntegrationMatrix" function, which integrates scATAC-seq with scRNA-seq. Specific marker genes used to identify individual cell types in scATAC-seq datasets of healthy $CD34^+$ hematopoietic cells and PBMCs are documented in supplementary table S2, Supplementary Material online.

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

mtDNA Variants Identified in Single-Cell ATAC-seq Datasets

Paired-end raw reads from each sample were aligned to the human reference genome (hg19) with Cell Ranger ATAC after adapter sequences were trimmed. First, the reads mapped to multiple sites or the nuclear genome, and duplicates were also removed. The remaining reads were realigned to correct the potential mapping errors around indels according to the process from GATK (McKenna et al. 2010). Bam files for each cell type were merged to identify germline mtDNA variants (bulk VAF $>$ 90%). Variants with VAF $>$ 90% shared among more than 90% cells were also considered germline mutations. Then mtDNA variants were called for each individual cell with VarScan2 (Koboldt et al. 2012) with "–min-var-freq 0.01," "min-depth 8," and "–min-reads2 2" at the beginning. To identify high confidence somatic variants in single cell, we applied the following filtering steps.

First, the germline mutations identified in the merged bam file were removed. Second, the following sites were explicitly removed because of the large numbers of homopolymers in the revised Cambridge Reference Sequence (rCRS) and sequencing errors in the reference genome (Ju et al. 2014): (1) misalignment due to ACCCCCCCTCCCCC (rCRS 302–315), including 302A/C, 309C/T, 311C/T, 312C/T, 313C/T, and 316G/C; (2) misalignment due to GCACACACACACC (rCRS 513–525), including 514C/A, 515A/G, 523A/C, and 524C/G; (3) misalignment due to 3107N in ACNTT (rCRS 3105–3109), including 3106C/A, 3109T/C, and 3110C/A. Third, sequencing errors can significantly affect the identification of somatic variants. Therefore, sequencing errors known to be associated with a high error rate according to Illumina NextSeq and sequence errors (G $\rightarrow$ T and C $\rightarrow$ A) from DNA damage were removed. Fourth, each mtDNA site count was required to be more than 20 reads (20$\times$), and the variant count more than two reads. Fifth, strand balance was required for confident somatic variants. For the given variant site, we required the reads mapped to the forward strand to be above 30% but below 70% of the total mapped reads for the variant allele. Variants that passed the multiple filter steps were merged from all individual cells as the final somatic variants. If the findings for a variant had sufficient confidence in any given cell, the VAF was recounted in all individual cells within the same cell type, with constraints of a minimum 8$\times$ depth. Cell with an average depth more than 10$\times$ were considered for further analysis.

Single-Cell RNA-seq Data Processing and Cell-Type Annotation

Downstream analysis of the scRNA-seq dataset was performed with Seurat (Stuart et al. 2019) (version 3.2.2; https://satijalab.org/seurat). The following bioinformatic analyses were performed in R software (version 3.6.0; https://www.r-project.org) with default settings unless otherwise stated. Cells with $<$200 or $>$2,500 detected genes or with $>$5% mtDNA were eliminated from further consideration. Normalization was applied with the MAGIC package (version 2.0.3) (van Dijk et al. 2018) by following the Seurat v3 workflow. We next calculated a subset of features that exhibited high cell-to-cell variability by using the "FindVariableFeatures" function and identified 2,000 specific features. Clusters were identified with the "Find-Neighbors" and "FindClusters" functions in Seurat with 45 principal components (PCs) and a resolution of 0.3. The results were annotated to include differential expression of cell type-specific marker genes. Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction was performed with the "RunUMAP" function in Seurat, with 45 PCs and other default parameters. The expression of cell type-specific marker genes in PBMCs and BMMCs is shown in supplementary table S3, Supplementary Material online. We referred to the information and classifications recorded in GSE139369 from the GEO database to guide our cell type annotations (supplementary table S3, Supplementary Material online).

Pseudotime Analysis

To construct single-cell differentiation trajectories with scRNA-seq data from HSCs to B cells, we performed a pseudotime analysis with the Monocle method (Trapnell et al. 2014; Qiu et al. 2017; Cao et al. 2019). First, we subdivided scRNA-seq data according to the annotated cell populations revealed by Seurat clustering analysis, according to the common pipeline (http://cole-trapnell-lab.github.io/monocle-release/monocle3/). Reclustering of selected cell populations was again performed with the "RunUMAP" function. Pseudotime analysis was conducted on these newly generated clusters with Monocle v3. We delineated expression patterns of G1/S phase-specific and mtDNA replication-related genes along a pseudotimeline. G1/S phase-specific genes were identified according to a previously annotated list (supplementary fig. S5*d*, Supplementary Material online) (Tirosh et al. 2016).

mtDNA Variants Identified from Single-Cell RNA-seq Data

mtDNA variants from single-cell RNA-seq data were processed in the same manner as mtDNA variants from scATAC-seq, with several modifications. Briefly, we used STAR (Dobin et al. 2013) to align reads to the human reference genome (hg19) and to obtain bam files. Germline mutations and mtDNA variants in individual cells were filtered and called in the same manner.

Allele Frequency Spectrum

The allele frequency (heteroplasmic ratio) of each mutation was calculated in each cell and the number of mutations that fall in each frequency bin (from 0 to 1) was counted for each cell type. Somatic mutations that arose in the early development stage, which had been fixed in

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

the progenitor cells, were further excluded for the AFS analysis in the mtscATAC-seq on BMMCs.

Annotation of Mitochondria DNA Mutations and Calculation of NS/SY Mutation Rates (*dN*/*dS*)

The mitochondrial variants were annotated with ANNOVAR (Wang et al. 2010). The annotated variants comprised mutations in loops, tRNA, rRNA, and mRNA coding regions, including NS and SY substitutions according to the variant location (supplementary figs. S8 and S9, Supplementary Material online). Coding sequences (CDS) within the mitochondrial genome were evaluated with Phylogenetic Analysis of Maximum Likelihood to identify all possible SY (defined as *S*) and NS (defined as *N*) substitutions in the human mitochondrial genome (Yang 2007). On the basis of ANNOVAR's annotations, we identified all observed SY (defined as *s*) and NS substitutions (defined as *n*). The NS mutation rate (dN) = n/N and the SY mutation rate (dS) = s/S, responses to positive, neutral, or negative selection pressure, can be determined by the dN/dS ratio.

Computational Modeling of the Mitochondrial Genetic Bottleneck

We used the Wright-Fisher model from population genetics to depict the accumulation of mutations and the dynamic frequency of heteroplasmic alleles in mtDNA during lymphoid cell divisions. The Wright-Fisher model assumes discrete generations and random sampling of individuals from the current generation without replacement by reproduction in the following generation. This model has been widely used to model the mtDNA population dynamics in both germline cells and somatic cells, including those that are neoplastic (Coller et al. 2001; Wilton et al. 2018). Because normal somatic cells typically contain 100–1,000 copies of mtDNA, we used $n = 500$ as the baseline copy number in our model (O'Hara et al. 2019). Results from the scATAC dataset revealed that the relative copy number of mtDNA in NK cells was ~60% that detected in B or T cells (fig. 3*a*); thus, 300 (500×0.6) was used as the baseline mtDNA copy number for the NK lymphocyte cohort. We modeled the lymphoid development from LMPP cells, which are the common progenitor cells for all lymphocytes, B, T, and NK cells. To model the dilution-based genetic bottleneck, we introduced a dilution rate α, which denotes the fraction of mtDNA molecules in each cell that undergo replication within a single-cell cycle, and T_d, which denotes the time of the diluting process. After T_d cell divisions from LMPP, the mtDNA copy number in each cell type rapidly recovers to the baseline level. The minimal mtDNA copy number through the bottleneck can be computed by:

$$N_b = N_0 \alpha^{T_d} \quad (1)$$

where N_0 is the initial number of mtDNA copies. The total number of cell divisions required for the transition from an LMPP to a mature lymphocyte is denoted T_a. The mutation rate at each site within the mitochondrial genome per cell division is denoted u, which has been estimated to be 10^{-8} to 10^{-7} mutations per site for somatic cells (Coller et al. 2001; Cabrera 2021). Thus, the mutation rate for the entire mitochondrial genome during each cell division event will be $u = \mu \times L$, where $L = 16{,}569$ base pairs (bp), representing the number of potential sites within the mtDNA length.

During each cell division, the number of somatic mutations acquired per mitochondrial genome follows a Poisson distribution with a mean of u. Thus, the probability that k mutations occurred in each cell division is as follows:

$$P(x = k) = \frac{u^k e^{-u}}{k!} \quad (2)$$

Computational Inference of Parameters by ABC

We used the framework of ABC for parameter inference in our computational model of somatic mtDNA population dynamics on the basis of the dilution rate α, the dilution time course T_d and the total number of cell divisions T_a. The minimal mtDNA copy number in each cell can be computed as described by equation (1) when values for α and T_d are available. The prior uniform distributions used for sampling α, T_d, and T_a, were, $T_d \sim U(0,\ 30)$ and $T_a \sim U(10,\ 40)$. To avoid extinction (i.e., a minimal mtDNA copy number of 0), only the sampled parameter values ensuring $N_b\ (= N_0 \alpha^{T_d}) > 10$ were retained. We used a version of ABC based on the acceptance–rejection algorithm (Tavare et al. 1997) to estimate posterior probability distributions for the parameters of interest (i.e., $\boldsymbol{\theta}$, T_d). We used 19 summary statistics ($\boldsymbol{S}$), which included the mtDNA mutation count in each VAF with a step size of 0.05 from VAF = 0.05 to 1 to fit the simulated to the observed data. The ABC version of rejection sampling is as follows:

1) Sample parameters θ' from the prior distribution $\pi(\boldsymbol{\theta})$
2) Simulate data ($\boldsymbol{D}'$) with the sampled parameters (θ') and summarize $\boldsymbol{D}'$ as summary statistics ($\boldsymbol{S}'$)
3) Accept θ' if $d(\boldsymbol{S}', \boldsymbol{S}) < \epsilon$, for a given tolerance rate ϵ, where $d(\boldsymbol{S}', \boldsymbol{S})$ is a measure of the Euclidean distance between $\boldsymbol{S}'$ and $\boldsymbol{S}$
4) Return to step 1.

With this scheme, we approximated the posterior distribution by $P(\boldsymbol{\theta} |\ d(\boldsymbol{S}', \boldsymbol{S}) < \epsilon)$. We used a common variation in ABC (Beaumont et al. 2002; Zhao et al. 2014) in which, rather than using a fixed threshold, ϵ, we sorted all calculated K distances by $d(\boldsymbol{S}', \boldsymbol{S})$ (see step 3 above) and accepted the θ' that generated the smallest $100 \times \eta$ percentage distances. We used $K = 10^6$ and $\eta = 0.001$ so that the posterior distribution was composed of $10^6 \times 0.001 = 1{,}000$ data points. We ran the ABC inference procedures for two mutation rates ($\mu = 10^{-8}$ and 10^{-7}) and

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

performed model selection (supplementary fig. S6, Supplementary Material online). The mutation rate $\mu = 10^{-7}$ fitted the data better in all cell types and thus was used for the computational inference. The ABC procedure was performed with the R package *abc* (Csilléry et al. 2012).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Weiwei Zhai (Institute of Zoology, Chinese Academy of Sciences) for constructive comments and suggestions on the manuscript. This work was supported by the National Key R&D Program of China (2021YFA1102100 to J.X.), the National Natural Science Foundation of China (32070644) to J.X., the Guangdong Basic and Applied Basic Research Foundation (2019A1515110387, 2019B1515130004 to J.X. and 2021B1515020042 to Z.H.), NIH grant RM1-HG007735 to H.Y.C. and China Postdoctoral Science Foundation funding (2021M693303) to Z.L. H.Y.C. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions

J.X., Z.H, and H.Y.C. designed and conceived the study. Z.T., B.C., and W.Z. collected and analyzed the data. Z.H. and Z.L. designed the computational model and performed the simulations. J.X., Z.H., Z.T., and Z.L. wrote the manuscript with inputs from all authors. All authors read and approved the final manuscript.

Data Availability

Detailed information on data resources is provided in supplementary table S1, Supplementary Material online. Code used for single-cell data analysis and computational modeling are available at https://github.com/tangzhj/Bottleneck.

Conflict of Interest. H.Y.C. is a co-founder of Accent Therapeutics, Boundless Bio, Cartography Biosciences, and is an advisor to 10x Genomics, Arsenal Biosciences, and Spring Discovery. The other authors declare no conflicts of interest.

References

Ahlqvist KJ, Leoncini S, Pecorelli A, Wortmann SB, Ahola S, Forsström S, Guerranti R, De Felice C, Smeitink J, Ciccoli L, *et al.* 2015. MtDNA mutagenesis impairs elimination of mitochondria during erythroid maturation leading to enhanced erythrocyte destruction. *Nat Commun.* **6**:6494.

Andersion S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, *et al.* 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**:457–465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* **23**:147.

Angajala A, Lim S, Phillips JB, Kim J-H, Yates C, You Z, Tan M. 2018. Diverse roles of mitochondria in immune responses: novel insights into immuno-metabolism. *Front Immunol.* **9**:1605.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**:2025–2035.

Buck MD, O'Sullivan D, Klein Geltink RI, Curtis JD, Chang C-H, Sanin DE, Qiu J, Kretz O, Braas D, van der Windt GJW, *et al.* 2016. Mitochondrial dynamics controls T cell fate through metabolic programming. *Cell* **166**:63–76.

Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY, Greenleaf WJ. 2018. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. *Cell* **173**:1535–1548.e16.

Cabrera VM. 2021. Human molecular evolutionary rate, time dependency and transient polymorphism effects viewed through ancient and modern mitochondrial DNA genomes. *Sci Rep.* **11**:5036.

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, *et al.* 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**:496–502.

Chandel NS. 2014. Mitochondria as signaling organelles. *BMC Biol.* **12**:34.

Chao T, Wang H, Ho P-C. 2017. Mitochondrial control and guidance of cellular activities of T cells. *Front Immunol.* **8**:473.

Chatre L, Ricchetti M. 2013. Prevalent coordination of mitochondrial DNA transcription and initiation of replication with the cell cycle. *Nucleic Acids Res.* **41**:3068–3078.

Chen B, Shi Z, Chen Q, Shen X, Shibata D, Wen H, Wu C-I. 2019. Tumorigenesis as the paradigm of quasi-neutral molecular evolution. *Mol Biol Evol.* **36**:1430–1441.

Coller HA, Khrapko K, Bodyak ND, Nekhaeva E, Herrero-Jimenez P, Thilly WG. 2001. High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat Genet.* **28**:147–150.

Csilléry K, François O, Blum MGB. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol.* **3**:475–479.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21.

Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756.

Ghosh SS, Fahy E, Bodis-Wollner I, Sherman J, Howell N. 1996. Longitudinal study of a heteroplasmic 3460 Leber hereditary optic neuropathy family by multiplexed primer-extension analysis and nucleotide sequencing. *Am J Hum Genet.* **58**:325–334.

Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, Greenleaf WJ. 2021. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet.* **53**:403–411.

Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, Parks B, Gars E, Liedtke M, Zheng GXY, *et al.* 2019. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol.* **37**:1458–1465.

Guo X, Wang Y, Zhang R, Gu Z. 2020. STAMP: a multiplex sequencing method for simultaneous evaluation of mitochondrial DNA heteroplasmies and content. *NAR Genom Bioinform.* **2**:lqaa065.

Hagström E, Freyer C, Battersby BJ, Stewart JB, Larsson N-G. 2014. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucl Acids Res.* **42**:1111–1116.

Jenuth JP, Peterson AC, Fu K, Shoubridge EA. 1996. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet.* **14**:146–151.

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022

Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, Davies HR, Papaemmanuil E, Gundem G, Shlien A, *et al.* 2014. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife* **3**:e02935.

Kazak L, Reyes A, Holt IJ. 2012. Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nat Rev Mol Cell Biol*. **13**: 659–671.

Koboldt DC, Zhang Q, Larson DE, Shen D, Mclellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. **22**:568–576.

Koehler CM, Lindberg GL, Brown DR, Beitz DC, Freeman AE, Mayfield JE, Myers AM. 1991. Replacement of bovine mitochondrial DNA by a sequence variant within one generation. *Genetics* **129**:247–255.

Lagouge M, Larsson N-G. 2013. The role of mitochondrial DNA mutations and free radicals in disease and ageing. *J Intern Med*. **273**: 529–543.

Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, Chiang Z, Pelka K, Verboon JM, Luo W, Christian E, *et al.* 2021. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat Biotechnol*. **39**:451–461.

Lee JW, Su Y, Baloni P, Chen D, Pavlovitch-Bedzyk AJ, Yuan D, Duvvuri VR, Ng RH, Choi J, Xie J, *et al.* 2022. Integrated analysis of plasma and single immune cells uncovers metabolic changes in individuals with COVID-19. *Nat Biotechnol*. **40**:110–120.

Lee C, Yen K, Cohen P. 2013. Humanin: a harbinger of mitochondrial-derived peptides? *Trends Endocrinol Metab*. **24**:222–228.

Lemasters JJ. 2005. Selective mitochondrial autophagy, or mitophagy, as a targeted defense against oxidative stress, mitochondrial dysfunction, and aging. *Rejuvenation Res*. **8**:3–5.

Lisci M, Barton PR, Randzavola LO, Ma CY, Marchingo JM, Cantrell DA, Paupe V, Prudent J, Stinchcombe JC, Griffiths GM, *et al.* 2021. Mitochondrial translation is required for sustained killing by cytotoxic T cells. *Science* **374**:eabe9977.

Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, Procaccio V, Wallace DC. 2013. mtDNA variation and analysis using Mitomap and Mitomaster. *Curr Protoc Bioinformatics* **44**:1.23.1–26.

Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A, *et al.* 2019. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**:1325–1339.e22.

Maciver NJ, Michalek RD, Rathmell JC. 2013. Metabolic regulation of T lymphocytes. *Annu Rev Immunol*. **31**:259–283.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**:1297–1303.

Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, Luo W, Huang T-S, Yeung BZ, Papalexi E, *et al.* 2021. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol*. **39**:1246–1258.

Miragaia RJ, Gomes T, Chomka A, Jardine L, Riedel A, Hegazy AN, Whibley N, Tucci A, Chen X, Lindeman I, *et al.* 2019. Single-cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *Immunity* **50**:493–504.e7.

O'Hara R, Tedone E, Ludlow A, Huang E, Arosio B, Mari D, Shay JW. 2019. Quantitative mitochondrial DNA copy number determination using droplet digital PCR with single-cell resolution. *Genome Res*. **29**:1878–1888.

Park J, Kang E, Kang S, Kim D, Kim D, Park SJ, Jhang WK. 2021. Mitochondrial gene mutations in pediatric septic shock. *Pediatr Res*. **90**:1016–1022.

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**:979–982.

Radsak K, Schutz E. 1978. Changes of mitochondrial DNA polymerase-gamma activity in synchronized mouse cell cultures. *Eur J Biochem*. **89**:3–9.

Sasaki T, Sato Y, Higashiyama T, Sasaki N. 2017. Live imaging reveals the dynamics and regulation of mitochondrial nucleoids during the cell cycle in Fucci2-HeLa cells. *Sci Rep*. **7**:11257.

Satpathy AT, Saligrama N, Buenrostro JD, Wei Y, Wu B, Rubin AJ, Granja JM, Lareau CA, Li R, Qi Y, *et al.* 2018. Transcript-indexed ATAC-seq for precision immune profiling. *Nat Med*. **24**:580–590.

Scheibye-Knudsen M, Fang EF, Croteau DL, Wilson DM 3rd, Bohr VA. 2015. Protecting the mitochondrial powerhouse. *Trends Cell Biol*. **25**:158–170.

Sena LA, Li S, Jairaman A, Prakriya M, Ezponda T, Hildeman DA, Wang C-R, Schumacker PT, Licht JD, Perlman H, *et al.* 2013. Mitochondria are required for antigen-specific T cell activation through reactive oxygen species signaling. *Immunity* **38**:225–236.

Starr TK, Jameson SC, Hogquist KA. 2003. Positive and negative selection of T cells. *Annu Rev Immunol*. **21**:139–176.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177**:1888–1902.e21.

Tan H, Yang K, Li Y, Shaw TI, Wang Y, Blanco DB, Wang X, Cho J-H, Wang H, Rankin S, *et al.* 2017. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity* **46**:488–503.

Tavare S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145**:505–518.

Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, *et al.* 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**:189–196.

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. **32**:381–386.

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D, *et al.* 2018. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**:716–729.e27.

Walker MA, Lareau CA, Ludwig LS, Karaa A, Sankaran VG, Regev A, Mootha VK. 2020. Purifying selection against pathogenic mitochondrial DNA in human T cells. *N Engl J Med*. **383**:1556–1563.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. **38**:e164.

Waters LR, Ahsan FM, Wolf DM, Shirihai O, Teitell MA. 2018. Initial B cell activation induces metabolic reprogramming and mitochondrial remodeling. *iScience* **5**:99–109.

West AP, Shadel GS. 2017. Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat Rev Immunol*. **17**: 363–375.

Wilton PR, Zaidi A, Makova K, Nielsen R. 2018. A population phylogenetic view of mitochondrial heteroplasmy. *Genetics* **208**:1261–1274.

Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, Majeti R, Chang HY. 2019. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* **8**:e45105.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. **24**:1586–1591.

Ye K, Lu J, Ma F, Keinan A, Gu Z. 2014. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci U S A*. **111**:10654–10659.

Yu YR, Imrichova H, Wang H, Chao T, Xiao Z, Gao M, Rincon-Restrepo M, Franco F, Genolet R, Cheng W-C, *et al.* 2020. Disturbed mitochondrial dynamics in CD8(+) TILs reinforce T cell exhaustion. *Nat Immunol*. **21**:1540–1551.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, *et al.* 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. **9**:R137.

Zhao J, Siegmund KD, Shibata D *et al.* 2014. Ancestral inference in tumors: how much can we know? *J Theor Biol*. **359**:136–145.

Downloaded from https://academic.oup.com/mbe/article/39/5/msac090/6585183 by guest on 16 May 2022